MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

DTIC FILE COPY

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 22000.8-LS	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) GENERAL POTENTIAL SURFACES AND NEURAL NETWORKS		5. TYPE OF REPORT & PERIOD COVERED TECHNICAL REPORT
		6. PERFORMING ORG. REPORT NUMBER TR-42
7. AUTHOR(s) AMIR DEMBO OFER ZEITOUNI		8. CONTRACT OR GRANT NUMBER(s) DAAG-29-84-K-0202 N00014-86-K-0041
9. PERFORMING ORGANIZATION NAME AND ADDRESS CENTER FOR NEURAL SCIENCE BROWN UNIVERSITY PROVIDENCE, RHODE ISLAND 02912		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office Post Office Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE June 24, 1987
		13. NUMBER OF PAGES 28 pages
14. MONITORING AGENCY NAME & ADDRESS(If different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

DTIC
ELECTE
SEP 17 1987
D

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, If different from Report)

NA

Superseded
AD-A181933

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Generalized Potential System
Hopfield Model
Associative Memory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Investigating Hopfield's model of associative memory implementation by a neural network, led to a generalized potential system with a much superior performance as an associative memory. In particular, there are no spurious memories, and any set of desired points can be stored, with unlimited capacity (in the continuous time and real space version of the model). There are no limit cycles in this system, and the size of all basins of attraction can reach up to half the distance between the stored points, by proper choice of the design parameters.

DD FORM 1473 EDITION OF 1 NOV 65 IS OBSOLETE   8
1 JAN 73

AD-A184 484

20. ABSTRACT CONTINUED

A discrete time version with its state space being the unit hypercube is also derived, and admits superior properties compared to the corresponding Hopfield network. In particular the capacity of any system of N neurons, with a fixed desired size of basins of attractions, is exponentially growing with N and is asympototically optimal in the information theory sense. The computational complexity of this model is slightly larger than that of the Hopfield memory, but of the same order.

The results are derived under an axiomatic approach which determines the desired properties and shows that the above mentioned model is the only one to achieve them.

Accesion For		
NTIS CRA&I		V
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1		

9

GENERAL POTENTIAL SURFACES AND NEURAL NETWORKS

by

Amir Dembo

and

Ofer Zeitouni

Lefschetz Center for Dynamical Systems
Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

January 1987

87   9   9 138

GENERAL POTENTIAL SURFACES AND NEURAL NETWORKS

Amir Dembo and Ofer Zeitouni

Division of Applied Mathematics
Brown University
Providence, Rhode Island 02912

ABSTRACT

Investigating Hopfield's model of associative memory implementation by a
neural network, led to a generalized potential system with a much superior per-
formance as an associative memory. In particular, there are no spurious memories,
and any set of desired points can be stored, with unlimited capacity (in the con-
tinuous time and real space version of the model). There are no limit cycles
in this system, and the size of all basins of attraction can reach up to half
the distance between the stored points, by proper choice of the design parameters.

A discrete time version with its state space being the unit hypercube
is also derived, and admits superior properties compared to the corresponding
Hopfield network. In particular the capacity of any system of N neurons, with a
fixed desired size of basins of attractions, is exponentially growing with N and
is asymptotically optimal in the information theory sense. The computational
complexity of this model is slightly larger than that of the Hopfield memory,
but of the same order.

The results are derived under an axiomatic approach which determines the
desired properties and shows that the above mentioned model is the only one
to achieve them.

INTRODUCTION

J. Hopfield's suggestion of a neural network model for associative memories in [1], arose the interest of many scientists and led to an effort of mathematically analysing its properties [2-12]. It is simple to implement this model, but hard to intuitively capture its properties, and even harder to rigouriously analyze its performance as an associative memory. This is perhaps the main reason for its attracting such interest.

The yet partial analysis done on this preliminary model revealed the following major disadvantages:

(a) There are many spourious memories generated at unexpected places (c.f., [3,7,8,9,11]), which attract a major part of the inputs.

(b) The capacity of the various versions of this model is bounded by N (the number of neurons), (c.f., [3,5,6,7,8,10]), which is quite a poor capacity compared to the Information Theory bounds on error correcting codes. Some suggestions as to how this bound can be enlarged appear in [12] but supply only a partial answer.

(c) Not only that the capacity is limited, it is context dependent, i.e., there are very small sets of memories which cannot be stored in the original Hopfield model, and the shape of a basin of attraction depends on far away attractors (c.f., [6]).

This motivated another suggestion of a continuous-time model with evolution by N Ordinary Differential Equations (ODE's), (c.f., [13]). The new model is reported experimentally to have better performance, although it still suffers from some of the drawbacks of its ancestor. Furthermore, a rigourious analysis of the ODE version seems to be almost impossible.

In this work we take a different approach. We start by assuming only the generic form of our model for associate memory, and derive its structure and properties out of a set of assumptions on the system.

The basic model we have in mind is of a set of memories ("particles", in classical mechanics or electrostatics, of specified "charge/mass"), which in the simplest case are located in the location of the desired memories. In general we allow for "spread out" charges (i.e., charge/mass densities instead of δ-functions). To each such "particle" α, we therefore associate its "charge density" μ_α. In general (unlike in classical mechanics), we allow for various "types" of particles (memories), i.e., the potential associated with each particle may be different (and this will affect the basins of attraction' shape). Indeed, the following three properties would be desirable for an associative memory model:

(P1) The system should be invariant to translations and rotations of the coordinates.

(P2) The system should be linear w.r.t. adding particles in the sense that the potential of two particles should be the sum of the potentials induced by the individual particles (i.e., we do not allow inter-particles interaction; see however, the discussion in Section IV).

(P3) Particle locations are the only possible sites of stable memory locations.

In order to state our results, we need to define exactly our system and describe how we build a memory out of the desired specifications.

In what follows, we take $\mathbb{R}^N$ to be our state space, and use first order, potential type ODE's to avoid the kind of "kinetic equilibria" one finds in second order equations, i.e., the equations of motion are:

$$\dot{x} = -\nabla V(x) \tag{1}$$

where $V(x)$ is our potential and ∇ stands for the gradient operator. Since we want to allow for various particle types, let us define a "type-space" A. A may be finite, countable, or even non-discrete, but clearly A is smaller than $\mathbb{R}^N$, and therefore finite dimensional. We assume that A is a measurable space so that integration over A is well defined. The specific examples for A that we have in mind are a finite, discrete set $\{1,2,\ldots,K\}$ or $\mathbb{R}^N$ itself.

Our memory building process is defined as a transformation $V(x) \overset{\Delta}{=} T(\mu(\cdot,\cdot))$

where $\qquad\qquad\qquad\qquad \mu \in M(\mathbb{R}^N \times A),$

and $M(\mathbb{R}^N \times A)$ stands for the space of measures over $\mathbb{R}^N \times A$ (which is clearly a linear space). For example, assume we want a potential $V_{x_0}(x)$ for a single particle of type a ($a \in A$) located at x_0. Then

$$V_{x_0}(x) = T(1_{x_0} \times 1_a)$$

(P1) - (P3) now read

(A1) $T(\mu(\mathbb{R}^N \times A)) = V(x) \Rightarrow T(\mu(c \mathbb{R}^N + \eta) \times A)) = V(cx + \eta)$ where $\eta \in \mathbb{R}^N$, $c \in \mathbb{R}^{N \times N}$ is a nonsingular, orthogonal matrix, and $c \mathbb{R}^N + \eta$ is the c-rotation, η-translation of $\mathbb{R}^N$.

(A2) T is a linear transformation over M, i.e.,

$$T(\mu(\mathbb{R}^N \times A)) = \int_{\mathbb{R}^N \times A} f(x,y,\alpha)\mu(dy \times d\alpha) \qquad (2)$$

where $f(x,y,\alpha)$ is the kernel of the transformation (Green's function) and we assume that $f(\cdot,\cdot,\cdot)$ is such that (2) makes sense.

(A3) Let $V(x) = T(\mu(\mathbb{R}^N \times A))$. Let $D = \{\overline{(x,\alpha)} \mid \mu(dx,\alpha) \neq 0\}$, where $\overline{\{\cdot\}}$ denotes the closure of a set. Then $V(x)$ does not possess minima outside of $D|_{\mathbb{R}^N}$, where $D|_{\mathbb{R}^N}$ is the restriction of D to $\mathbb{R}^N$.

In addition, we assume throughout the necessary smoothness and growth conditions on $f(x,y,\alpha)$ and its derivatives (w.r.t. x, y). In particular, we assume that if $f(x_0,y_0,\alpha_0)$ is finite then $f(x,y,\alpha_0)$ is twice continuously differentiable w.r.t. x at (x,y_0), $\forall x \neq y_0$, with integrable derivatives w.r.t. $\mu(\cdot,\cdot)$.

Our first result, which is proven in the appendix, is the following structure theorem:

Theorem 1. $V(\cdot)$ satisfy (A1) - (A3) in a universal manner w.r.t. every $\mu \in M$ iff it is of the structure:

$$V(x) = \int_{\mathbb{R}^N \times A} f_\alpha(||x-\eta||^2)\mu(d\eta \times d\alpha) \tag{3}$$

where $\forall \alpha \in A$, $f_\alpha(||x-\eta||^2)$ is a sub-harmonic function on $\mathbb{R}^N - \{\eta\}$, i.e.,

$$\forall d > 0 \qquad f_\alpha''(d)d + \frac{N}{2} f_\alpha'(d) \leq 0. \tag{4}$$

Solving (4) under the assumptions that $f_\alpha'(d_0) \neq 0$, for some $d_0 > 0$, and adding the proper constant so that $f_\alpha(d_0) = - \dfrac{f_\alpha'(d_0)d_0}{(\frac{N}{2} - 1)}$, (with the details given in the Appendix), we obtain for $N \geq 3$:

Lemma 1. Any solution of (4) can possess at most one local maxima (and no local minima), for $d \in (0,\infty)$, and satisfies:

$$f_\alpha(d) \geq f_\alpha(d_0)\left(\frac{d}{d_0}\right)^{-(\frac{N}{2} -1)} , \quad \text{for } d \geq d_0 \tag{5a}$$

$$f_\alpha(d) \leq f_\alpha(d_0)\left(\frac{d}{d_0}\right)^{-(\frac{N}{2} -1)} , \quad \text{for } d \leq d_0. \tag{5b}$$

Remarks. 1. By local minima (maxima) we mean a strict minima (maxima). For strictly sub-harmonic functions, we can assure the also non-existence of non-strict extrema.

2. Equality in equations (5) for every $d > 0$ implies equality in (4), which implies that $f_\alpha(||x-\eta||^2)$ is an harmonic function on $\mathbb{R}^N - \{\eta\}$. Whenever this holds $\forall \alpha \in A$, not only there are no strict local minima of $V(x)$ outside $D|_{\mathbb{R}}$, but there are also no strict local maxima of $V(\cdot)$ at those points.

For this particular case (where for simplicity we use $d_0 = 1$, w.l.o.g.):

$$V(x) = \int_{\eta \in \mathbb{R}^N} ||x-\eta||^{-(N-2)} \overline{\mu}(d\eta) \tag{6}$$

where

$$\overline{\mu}(\cdot) \triangleq \int_{\alpha \in A} f_\alpha(1) \mu(\cdot, d\alpha),$$

is a <u>signed measure</u> on $\mathbb{R}^N$.

The derivation of the representation of $V(\cdot)$ was done under the most general condition. Usually however, one is interested in an associative memory with a discrete number of memories, possibly of different types. Let therefore $A = \{1,\ldots,K\}$, where K is the number of particles, and let the α-th memory be located at $u^{(\alpha)} \in \mathbb{R}^N$, i.e., $\mu(\eta,\alpha) = 1_{\eta=u^{(\alpha)}} \times 1_{\alpha \leq K}$. We concentrate from now on on this class of systems, which is represented by

$$V(x) = \sum_{\alpha \in A} f_\alpha(||x-u^{(\alpha)}||^2) \tag{7}$$

For example, when we combine (6) and (7), we obtain:

$$V(x) = \sum_{\alpha \in A} \frac{f_\alpha(1)}{||x-u^{(\alpha)}||^{(N-2)}} \tag{8}$$

which is exactly the Electrostatical potential in the particular case of $N = 3$.

For any value of $N \geq 3$, and $V(\cdot)$ given by (7), with $f_\alpha(\cdot)$ satisfying (4), we distinguish between two types of memories:

(A) Attractive memories, with $f_\alpha(\cdot)$ monotonically increasing at least at some $d_0 < \infty$, in which case $\lim_{d \to 0} f_\alpha(d) = -\infty$ (from (5b)), and from (5a) we note

that (at least when $f_\alpha(\cdot)$ is non-decreasing everywhere), the potential induced by the memory at $u^{(\alpha)}$ should approach rapidly its limiting value as (d/d_0) increases. For those memories:

$$\lim_{x \to u^{(\alpha)}} V(x) = -\infty \qquad (9a)$$

so that they are the global (and also local) minima of the potential function, and correspond to relatively short-range interactions.

(B) Repulsive memories, with $f_\alpha(\cdot)$ monotonically non-increasing, in which case $\lim_{d \to 0} f_\alpha(d) > -\infty$.

In that case, $u^{(\alpha)}$ is not a strict local minima of $V(\cdot)$, and two behaviors are possible:

(B1) $\lim_{d \to 0} f_\alpha(d) = \infty$, $\lim_{d \to \infty} f_\alpha(d) = const$. An example is the electrostatic force of a <u>negative</u> particle (instead of the positive one in (A)). Those are relatively strong, short-range interactions, and are of interest if one wishes to "avoid" specific locations (as now $u^{(\alpha)}$ is a strict global maxima of $V(\cdot)$).

(B2) $\lim_{d \to 0} f_\alpha(d) = 0$, (and usually, $\lim_{d \to \infty} f_\alpha(d) = -\infty$). An example is $f(d) = -d$ ("repulsive spring"), and those forces are weak in the short range but strong in the long range. We do not use those kind of repulsive memories in the sequel.

In particular, for the Electrostatical form of the potential given in (8), $V(\cdot)$ is an harmonic function outside $\{u^{(\alpha)}\}_{\alpha \in A}$, thus possess all its local minima in the attractive memories, and all its local maxima in repulsive memories of type (B1). Thus, we can store in the same system two kinds of objects. While the recall process using (1), will give rise to objects of type (A), the same recall process with $-V(\cdot)$ instead of $V(\cdot)$ will give rise to objects of type (B1).

The potentials given by (4) and (7), possess the major property one expects from an associative memory. The desired memories are arbitrarily chosen, with their recall being guaranteed, and their number and distribution unrestricted. Furthermore, our assumption (A3), together with the properties of the potential

type ODE's in (1) (c.f. [14]) guarantee that except for a set of measure zero of saddle points, every initial probe x(0) will converge to a desired memory $u^{(\alpha)}$ of type (A).

Our assumptions, (A1) and (A2), made the mathematical analysis tractable. When some are omitted the class of potentials with property (A3) is enlarged. For example without (A1) we obtain (3) with $f_\alpha(x,\eta)$ instead of $f_\alpha(||x-\eta||^2)$, and (4) is replaced by:

$$\forall \eta \in \mathbb{R}^N \quad \forall x \in \mathbb{R}^N - \{\eta\} \qquad \Delta_x f_\alpha(x,\eta) \leq 0 \tag{10}$$

where Δ_x stands for the Laplacian operator w.r.t. x. This corresponds to a "non-homogeneous" state space, but complicates the mathematical analysis.

To compare our class of "neural networks" with the model of [1], as well as the Information Theory bounds on error correcting codes, we derive the discrete-time finite state space analogue of the evolution (1).

Consider the state space as the unit hypercube in $\mathbb{R}^N$, to be denoted by H^N. For any potential function $V(x)$, the relaxation algorithm is (in the spirit of [12]):

(A) According to some predetermined probability measure peak a point $y \in H^N$ having Hamming distance one from the current state $x \in H^N$.

(B) If $V(y) < V(x)$, then the new state will be y, otherwise it remains x. In both cases return to step (A).

As shown in [12], for any $V(\cdot)$ and $x^{(0)}$, this algorithm converges to a fixed point in H^N. For any practical memory of this type, $\{u^{(\alpha)}\}_{\alpha \in A} \subset H^N$, and thus A is a finite set with $K = |A| \leq 2^N$.

Whereas the class of memories suggested here is of the form:

$$V(x) = \sum_{i=1}^{K} f_i(||x-u^{(i)}||^2) \tag{11}$$

with $f_i(\cdot)$ satisfying (4), the model suggested in [1] corresponds to (11) with: $f_i(d) = \frac{1}{2}[N - (N - \frac{1}{2}d)^2]$ which does not satisfy (4).

Note that the more complex versions of this model (c.f., [5,6,8]), does not satisfy assumptions (A1), (A2) at all.

In the next section we analyze the continuous-time model (the ODE's evolution), in terms of the basins of attraction, and convergence rate analysis. In section III, the discrete time version is analyzed. The capacity (K), is related to the error correction capability, and compared with known results on Hopfield's model. The last section is devoted to rough complexity analysis for both models, as well as a comparison with the classical Hamming classifier (using minimal distance search), and to possible generalizations of (1) which allow for more complicated tasks as clustering, supervised learning, etc.

II. BASINS OF ATTRACTION AND CONVERGENCE RATE

The potential given by (7) usually allows for an infinite number of distinct stable states (memories), thus having infinite capacity. This however does not reveal the shape of the basins of attractions of these memories.

For analyzing the performance of the system in (7) as an associative memory, assume the simplified assumptions that $f_\alpha(\cdot)$ is a monotonically nondecreasing function, independent of α and that $||u^{(\alpha)} - u^{(\beta)}|| \geq 1$, for every $\alpha \neq \beta \in A$. We shall investigate the value of:

$$\varepsilon_{max} \triangleq \min_{\{u^{(\alpha)}\}_{\alpha \in A}} \left\{ \max_\rho \{\rho; \text{ s.t. } ||x(0)-u^{(\alpha)}|| \leq \rho \text{ implies } x(t) \underset{t\to\infty}{\to} u^{(\alpha)}\} \right\}. \qquad (12)$$

It is clear from symmetry arguments that $\varepsilon_{max} \leq 1/2$ (where the outer minimization is over all possible positions of $\{u^{(\alpha)}\}_{\alpha \in A}$ in $\mathbb{R}^N$).

Our aim is to show that a proper choice of $f(\cdot)$ (which is sub-harmonic) will lead to ε_{max} as close to 1/2 as desired, thus the "maximal" basin of attractions can be guaranteed.

The following lower bound on ε_{max} is derived in the Appendix, by bounding the maximal contribution of farther away particles to forces in the boundary of the ε_{max} sphere around $u^{(\alpha)}$:

Lemma 2. (a). ε_{max} is larger than any value of $r < 1$ satisfying:

$$f'(r^2)r \geq \int_1^\infty - \frac{d}{dt} \{(t-r)f'((t-r)^2)\}(2t+1)^N dt. \qquad (13)$$

(b). This is a tight bound in the sense that whenever the r.h.s. of (13) diverges then $\varepsilon_{max} = 0$.

We shall now restrict our attention to $f(d) = -k(\frac{d}{d_0})^{-m}$ with $m \geq (\frac{N}{2} - 1)$, integer, (this guarantees that $f(\cdot)$ satisfies equation (4)).

The r.h.s. of (13) is finite iff $m > \frac{N}{2} - \frac{1}{2}$, and then, for integer m, (13) is exactly:

$$(km \ d_0^m) r^{-(2m+1)} \geq (km \ d_0^m) 3^N (1-r)^{-(2m+1)} \sum_{k=0}^{N} \frac{\binom{N}{k}}{\binom{2m}{k}} \left[\frac{2}{3}(1-r)\right]^k \tag{14}$$

which leads to

$$\frac{1}{2} \geq \varepsilon_{max}(m,N) \geq \left\{1 + \left[\frac{1}{2} 3^{(N+1)}\right]^{1/(2m+1)}\right\}^{-1} \tag{15}$$

so, for any N, $\lim_{m\to\infty} \{\varepsilon_{max}(m,N)\} = 1/2$, and for any $m = \frac{1}{2}(k(N+1) - 1)$ with $k \geq 1$ fixed, $\lim_{N\to\infty} \{\varepsilon_{max}(\frac{1}{2}(k(N+1) - 1),N)\} = 1/(1 + \sqrt[k]{3})$. So even for $m = N/2$, $\varepsilon_{max}(m,N) \sim \frac{1}{4}$, for N large enough.

Remarks. 1. If the $\{u^{(\alpha)}\}_{\alpha \in A}$ are restricted to be contained in a sphere of radius ρ in $\mathbb{R}^N$, then the integral in the r.h.s. of (13) will have upper limit 2ρ, and the additional term $(2\rho-r)f'((2\rho-r)^2)(4\rho+1)^N$ would be added there. It is thus finite for every value of m (including the harmonic case $m = \frac{N}{2} - 1$), implying $\varepsilon_{max} > 0$ under this restriction.

2. For $N \to \infty$, and fixed k, the $1/(1 + \sqrt[k]{3})$ behaviour is maintained even when we consider only memories $\{u^{(\alpha)}\}_{\alpha \in A}$ which are contained in the unit sphere ($\rho = 1$), as a refinement of the arguments of Lemma 2 shows.

As for the rate of convergence, it is easy to verify that for a large value of m, and $x(0)$ far from all the $\{u^{(\alpha)}\}_{\alpha \in A}$, it will take a long time for the evolution in (1), before $x(t)$ will be near one of the $\{u^{(\alpha)}\}_{\alpha \in A}$. However (as we prove in the Appendix):

Lemma 3. Let Q be any closed set in $\mathbb{R}^N$ whose interior includes the convex hull of $\{u^{(\alpha)}\}_{\alpha \in A} \cup \{\overline{u}\}$, where $\overline{u}$ is an arbitrary point in $\mathbb{R}^N$ (possibly within the convex hull of $\{u^{(\alpha)}\}_{\alpha \in A}$). Then, adding $1_{x \notin Q} \ g(||x-\overline{u}||^2)$ to $V(x)$, where $g(x)$

is any nondecreasing, differentiable function will not disturb (A3), nor create additional fixed points to (1), provided all the $f_\alpha(\cdot)$ which compose $V(x)$ are monotonically increasing.

Therefore, if for example $g(d) = d^\beta$ is added (with $\beta > 1$), then the convergence from $x(0)$ at infinity to a point with squared distance d_0 from the points $\{u^{(\alpha)}\}_{\alpha \in A}$ and $\bar{u}$ (with d_0 much larger then the squared distances between these $|A| + 1$ points), takes the time $T \sim d_0^{-(\beta-1)}/4\beta(\beta-1)$. Thus, by using β, large enough, convergence from infinity to ∂Q (the boundary of Q) can take arbitrarily small time.

Global investigation of the rate of convergence inside the convex hull of $\{u^{(\alpha)}\}_{\alpha \in A}$ is quite cumbersome. Thus, let us restrict again the discussion to the case where $||u^{(\alpha)} - u^{(\beta)}|| \geq 1$, $f(d) = -k(\frac{d}{d_0})^{-m}$ (with $m \geq \frac{N}{2}$, is an integer). Furthermore, let $x(0)$ satisfy $||x(0) - u^{(\alpha)}|| \leq \theta\hat{\epsilon}(m,N)$ where $\theta < 1$, and $\hat{\epsilon}(m,N)$ is the lower bound on $\epsilon_{max}(m,N)$ given by the r.h.s. of (15).

We have seen already that for every $\theta < 1$, $x(t) \underset{t \to \infty}{\to} u^{(\alpha)}$, but (as we prove in the Appendix):

Lemma 4. Under the above conditions $x(T) = u^{(\alpha)}$, where:

$$T = \left(\frac{\theta\hat{\epsilon}(m,N)}{\sqrt{d_0}}\right)^{(2m+2)} \left(\frac{d_0}{2mk}\right) \left\{1 - \left[\frac{\theta(1 - \hat{\epsilon}(m,N))}{(1 - \theta\hat{\epsilon}(m,N))}\right]^{2m+1}\right\}^{-1}. \tag{16}$$

So that for m large enough, $\sqrt{d_0} = \hat{\epsilon}(m,N)$, and $k = d_0/2m$, we obtain $\log T \sim 2(m+1)\log \theta$. Again, by enlarging m while preserving θ fixed, T can become arbitrarily small.

To conclude - the "maximal" basins of attraction can be guaranteed by enlarging m (choosing strong, short range interactions), and this will also speed up the convergence within these basins of attraction (which is completed in an

arbitrary small <u>finite</u> time). The convergence from infinity to the neighborhood $\{u^{(\alpha)}\}_{\alpha \in A}$ can be fasten, without affecting all those properties, by the mechanism suggested in Lemma 3 (adding a long range field outside a proper set Q).

III. DISCRETE TIME EVOLUTION ON THE UNIT HYPERCUBE

Whereas the discrete time algorithm presented in the introduction uses $V(\cdot)$ which has no local minima outside $\{u^{(\alpha)}\}_{\alpha \in A}$, it might possess <u>fixed points</u> out of this set. The reason for that is the "rigidity" of the algorithm which might not allow descent in the gradient direction, due to the limited search for lower potential only in the Hamming distance one neighborhood of every $x \in H^N$.

We can, however, show that the proposed class of potentials is optimal according to the Information Theory bounds (as $N \to \infty$), and in particular can be used to design error correcting codes with positive rate (c.f., [15]).

Let us restrict the discussion to potentials of the form (11), with $f(d) = -d^{-m}$, $m \geq \frac{N}{2} - 1$. For simplicity of notation we use the normalized Hamming distance $||x-y|| \overset{\Delta}{=} \frac{1}{N} \sum_{i=1}^{N} |x_i - y_i|$, so that H^N is contained in the unit sphere, and assume that $\forall i \neq j$, $||u^{(i)} - u^{(j)}|| \geq 2\rho$, with $\frac{1}{2} \geq \rho > \frac{1}{N}$, fixed. Therefore, these code words $\{u^{(i)}\}_{i=1}^{k}$, can tolerate up to ρN errors in N coordinates.

As we prove in the Appendix, by bounding the total "force" the farther away particles apply on $x(0)$, we obtain:

<u>Theorem 2</u>. For every $x(0)$ such that $||x(0) - u^{(i)}|| \leq 2\theta\rho$, $\frac{1}{2} \geq \theta \geq 0$, and:

$$(K-1) \leq \left(\frac{1-\theta}{\theta}\right)^{2m} \left[\frac{1 - (1 + \frac{2}{N\rho})^{-2m}}{(1 - \frac{2}{N\rho})^{-2m} - 1}\right]. \tag{17}$$

(a). The discrete time algorithm will generate a sequence of states $\{x(n)\}_{n=1}$, such that $\forall n$, $||x(n) - u^{(i)}|| \leq ||x(n-1) - u^{(i)}||$ with equality iff $x(n-1) = x(n)$.

(b). There are exactly $\frac{N}{2}||x(0) - u^{(i)}||$ distinct states in this sequence, and if each coordinate has positive probability to be chosen as the updated coordinate, then $x(n)$ converges to $u^{(i)}$ with probability one.

Consider now $\rho > 0$, and $\theta < \frac{1}{2}$ fixed, with $m \geq \frac{N}{2}\left\{\log_{2+\epsilon}\left(\frac{1-\theta}{\theta}\right)\right\}^{-1}$ (where $\epsilon > 0$ is arbitrarily small). For this case, if N is large enough the r.h.s. of (17) is larger than 2^N, thus K is determined only by the bounds on the maximal number of points in H^N satisfying $\forall i \neq j$, $||u^{(i)} - u^{(j)}|| \geq 2\rho$, i.e., Information Theory asymptotic, sphere packing bounds for error correcting codes (c.f., [15]).

To conclude, Theorem 2 guarantees that for short range forces (i.e., $m(N)$ large enough), and large enough dimension, direct convergence (c.f., [3] for this definition) to the nearest code word is obtained, independently on the number of code words and their locations, provided that its initial distance is smaller than $\frac{1}{2} \min_{i \neq j} ||u^{(i)} - u^{(j)}||$.

For comparison, for the model of [1] which has "strong" forces, the maximal number of memories is bounded above by N even for $\underline{\theta = 0}$ (i.e., recall with no errors). Thus, this model has zero rate when referred to as an error correcting code (c.f., [16]). Even when it converges, the convergence time might grow exponentially with N, unlike the linear time guaranteed by Theorem 2 (when proper selection of the updated coordinate is done).

Although Theorem 2 was derived for the simplest case of equal desired basins of attraction, it can be generalized to other situations.

IV. LEARNING AND COMPLEXITY

A. Learning

The associative memory models presented in the preceeding two sections are capable of both storing information and recalling it. The analysis was restricted to the Euclidean state space, and the unit hypercube (equipped with the Hamming norm), merely for simplicity of presentation, and to enable comparison with Hopfield models (c.f., [1,13]).

When the state space is an arbitrary Riemannian manifold, the evolution (1) can be defined more abstractically as the potential ODE's on that manifold, with the gradient and Laplacian operators in (1) and (10) being defined on the manifold. As the maximum principle/Gauss theorem (c.f., [16]), which was the key to Theorem 1, is valid also on any Riemannian manifold, most of the results in this work can be extended to this more general context.

As for the discrete time version of the algorithm, it can be easily extended to any finite graph whose vertices are embedded in $\mathbb{R}^N$, (c.f., [12]).

The process of storage and recall of information described in this work does not involve any learning nor generalization (in the sense of [17]). It is also uncapable of creating periodical orbits (as done for example, in [4]). However, by further generalizing the kinematical laws, one can incorporate most of these phenomena.

For example, periodical orbits can be generated by modifying (1) to the more "classical" equation of motion:

$$\ddot{x} = -\frac{1}{m}\nabla V(x) - \mu\dot{x} \tag{18}$$

which for $N = 3$ is just the Newtonian motion of particle with mass m, in the field of the potential $V(\cdot)$, with viscosity coefficient μ.

Likewise, learning can be obtained by modifying the locations of $\{u^{(\alpha)}\}_{\alpha \in A}$ during the recall operation, either as a response to the distribution of the initial states $x(0)$, or to an external teaching procedure, or by adding inter-particles interactions. These modifications can be implemented within the evolution (1) (or (18)), by allowing the state $x(t)$ to be represented by a non-negligable particle, which apply <u>forces</u> on the given $\{u^{(\alpha)}\}_{\alpha \in A}$. Generalization (which is basically a spontaneous creation of clustering) is easily obtained once the $\{u^{(\alpha)}\}_{\alpha \in A}$ particles are allowed to apply forces one on the other, and change their locations.

Of course, in order to make all these remarks valid, goals should be defined rigorously, and mathematica/physical rules that will achieve them should be incorporated within this framework.

We conclude this subject by pointing out that we have shown that there is nothing special in Spin-Glass models, and other known models in phys and mathematics possess the "emergent collective computational abilities", onc are properly interpreted.

B. Complexity

Our proposed models have better performance than the models in [1,13], but what about the implementation complexity?

For comparison purposes we deal with three algorithms on H^N. The first one is the classical Hamming decoder w.r.t. to $\{u^{(i)}\}_{i=1}^{K} \subset H^N$. It involves the parallel computation of the K correlations $(u^{(i)}, x(0))$ (where $x(0) \in H^N$ as well), followed by a search for the maximal value, implemented in a tree structure. Thus, KN multiplications are needed together with $K \log K$ comparisons of pairs

of numbers, and the delay of the algorithm is $\log K + 1$ "unit" times (where comparison and multiplication are assumed equivalent throughout).

The second algorithm is the one suggested in [1]. Each iteration involves KN multiplications and N comparisons of pairs of numbers (since $K \leq N$ for this algorithm, as shown in [3,6,7,9,10]). The time delay however, is the number of iterations ("full sweeps", as all N coordinates are updated asynchronously), which is believed to be independent of K.

The last algorithm is the one suggested here in Eq. (11), with $f_i(d) = -d^{-m}$. It involves KN multiplications in each iteration for obtaining the d's. The operation of $f_i(\cdot)$ is quite simple once done by an analogue computer: One diode takes $\log d$, then multiplication by $(-m)$ is done, and at last a second diode computes $-\exp[-m(\log d)] = -d^{-m}$. So the overall complexity is again determined by the KN multipliers, and the time delay (in "full sweeps") is again a small constant as Theorem 2 implies.

Thus, for $K \leq N$, the new algorithm has the same complexity as Hopfield's scheme, and the classical Hamming decoder, with smaller time delay for the first two algorithms.

This result is true also for $K \sim 2^{h(\rho)N}$, but then the Hopfield model cannot be used, whereas the new algorithm has complexity which is linear in K, i.e., exponential in N. In this case it is better (in time delay) then the classical Hamming decoder, but does not admit the polynomial complexity of some of the special error correcting codes used in coding theory (c.f., [15]).

ACKNOWLEDGEMENT

APPENDIX

Proof of Theorem 1.

We use assumption (A1) for the case of atomic measures on $A \times \mathbb{R}^N$, i.e.,

$V(x) = f(x, \eta_0, \alpha_0)$. Consider first a translation of the coordinates, i.e.,

$x' = x + \Delta$, with proper translation of the atomic measure μ, i.e., $\eta_0' = \eta_0 + \Delta$.

As assumption (A1) implies that $V(x') = V(x)$, $f(x, \eta_0, \alpha_0) = f(x+\Delta, \eta_0+\Delta, \alpha_0)$ for

every $\eta_0, x, \Delta \in \mathbb{R}^N$ and every $\alpha_0 \in A$. Thus, $f(x, \eta_0, \alpha_0)$ depends only on $x - \eta_0$.

Repeating this argument for the case of rotation of the coordinates will prove

that $f(x, \eta_0, \alpha_0)$ depends only on $||x - \eta_0||^2$ for every $\alpha_0 \in A$.

Thus, the structural assumptions (A1) and (A2) impose that $V(\cdot)$ is of the

form given in (3).

We now assume that (4) is satisfied for every $\alpha \in A$. It is easily verified

that (4) is equivalent to:

$$\forall x \in \mathbb{R}^N - \{\eta\} \quad \Delta_x f_\alpha(||x-\eta||^2) \leq 0 \qquad (A.1)$$

where Δ_x is the Laplacian operator w.r.t. x. Equation (A.1) implies in view

of (3), that for every neighborhood $U(x)$ of $x \in \mathbb{R}^N$,

in which $\int_{\alpha \in A} \mu(\eta \times d\alpha)$ is identically zero,

$\Delta_x V(\cdot) \leq 0$ (in $U(x)$). In deriving this result we used the smoothness assumption

on V, together with integrability assumption on $\Delta_x f(\cdot)$ to allow for changing the

order of differentiation and integration. Suppose now that $f_\alpha(\cdot)$ satisfies (4)

$\forall \alpha \in A$, but there exists a local minima at $x_0 \in \mathbb{R}^N$ with $U(x_0)$ in which

$\int_{\alpha \in A} \mu(\eta \times d\alpha) = 0$. On $U(x_0)$, $\Delta_x V(\cdot) \leq 0$, so the maximum principle implies that the

minimum of $V(\cdot)$ in any closed subset of $U(x_0)$ is obtained on the boundary of

$U(x_0)$ (c.f., [18]). However, since x_0 is a local minima there exists a small, closed neighborhood around it such that $V(x_0) < \text{Inf } V(x)$ on that neighborhood, so contradiction is obtained.

Remark: We have shown that Eq. (4) guarantees that assumption (A3) holds, where we interprete as local minima only isolated points. Refinement of the above arguments leads to the elimination of constant surfaces of local minima, whenever strict inequality holds in (4).

To complete the "only if" part of Theorem 1, we assume that (4) does not hold for $\alpha_0 \in A$ and $d_0 > 0$, and consider the case of $V(\cdot)$ generated by (3) with $\mu(d\eta \times d\alpha)$ being an atomic measure on α_0 × a uniform measure on the sphere of radius $\sqrt{d_0}$ in $\mathbb{R}^N$. We shall prove that in this case there is a spourious local minima of $V(\cdot)$ at $x = 0$, which contradicts assumption (A3), as $d_0 > 0$.

At $x = 0$, $||x - \eta||^2 = d_0$ for every η on the sphere of radius $\sqrt{d_0}$, which implies that $\Delta_x V(\cdot) > 0$ at $x = 0$ (since (4) and therefore (A.1) does not hold there). The continuity of $\Delta_x V(\cdot)$ near $x = 0$ is imposed by our smoothness assumption, and guarantees that there is a spherical neighborhood $U(0)$ where $\Delta_x V(\cdot) > 0$. Since the measure μ is spherically symmetric, so is the potential $V(\cdot)$, (i.e., $V(x)$ depends only on $||x||$). We now apply the maximum principle on the concentric spheres contained in $U(0)$, and see that in any such sphere the maximum of $V(\cdot)$ is obtained only on the boundary. However, the spherical symmetry of $V(\cdot)$ implies it is constant on these boundaries; i.e., $x = 0$ is a local minima of $V(\cdot)$, as we claimed above. □

Proof of Lemma 1. Whenever $f'_\alpha(d) = 0$, (4) implies that $f''_\alpha(d) \leq 0$, so that $f'_\alpha(d)$ can cross the zero level only once in $d \in (0, \infty)$, and with $f''_\alpha(d) < 0$. Thus, solutions of (4) will possess at most one local maxima and no local

minima in $(0,\infty)$. It is easy to verify that (4) is equivalent to

$$f'_\alpha(d)d^{N/2} \quad \text{is monotonically non-increasing on } (0,\infty). \tag{A.2}$$

Thus:

$$f'_\alpha(\tilde{d}) \leq f'_\alpha(d_0)\left(\frac{d_0}{\tilde{d}}\right)^{N/2}, \quad \text{for } \infty > \tilde{d} \geq d_0 > 0 \tag{A.3a}$$

$$f'_\alpha(\tilde{d}) \geq f'_\alpha(d_0)\left(\frac{d_0}{\tilde{d}}\right)^{N/2}, \quad \text{for } 0 < \tilde{d} \leq d_0 < \infty. \tag{A.3b}$$

Integrating (A.3a) from d_0 to $d \geq d_0$, and using the condition $f_\alpha(d_0) = -\dfrac{f'_\alpha(d_0)d_0}{(\frac{N}{2} - 1)}$ will lead to (5a), whereas integrating (A.3b) from $d \leq d_0$ to d_0 will lead to (5b), for $N \geq 3$. Similar results can be obtained for $N = 1,2$, but are less interesting. □

Proof of Lemma 2. (a) Since $f(\cdot)$ is sub-harmonic (satisfies (4)), and monotonically nondecreasing ($f'(d) \geq 0$), it follows that $f(\cdot)$ atisfies (4) also for $N = 1$, i.e., that $f'(r^2)r$ is a monotonically nonincreasing function of $r \in \mathbb{R}_+$. For every $r > 0$ such that $||x-u^{(\alpha)}|| \leq r$ implies $(\nabla V, x-u^{(\alpha)}) \geq 0$, independently of the locations of the other $\{u^{(\alpha)}\}_{\alpha \in A}$ memories, the evolution (1) will monotonically decrease $||x-u^{(\alpha)}||$ until $x(t) = u^{(\alpha)}$. Thus, any such r is a lower bound on ϵ_{max}. In view of (7):

$$\frac{1}{2||x-u^{(\alpha)}||} (\nabla V, x-u^{(\alpha)}) = \sum_{\beta \in A} f'(||x-u^{(\beta)}||^2) \frac{(x-u^{(\beta)}, x-u^{(\alpha)})}{||x - u^{(\alpha)}||} \geq$$

$$\geq f'(||x-u^{(\alpha)}||^2)||x-u^{(\alpha)}|| - \sum_{\beta \neq \alpha} f'(||x-u^{(\beta)}||^2)||x-u^{(\beta)}|| \geq$$

$$\geq f'(||x-u^{(\alpha)}||^2)||x-u^{(\alpha)}|| - \sum_{\beta \neq \alpha} f'((||u^{(\alpha)}-u^{(\beta)}||-||x-u^{(\alpha)}||)^2) \cdot$$

$$\cdot (||u^{(\alpha)}-u^{(\beta)}||-||x-u^{(\alpha)}||) \geq$$

$$\geq f'(r^2)r - \sum_{n=0}^{\infty}(L_{1+(n+1)\epsilon} - L_{1+n\epsilon})f'(((1+n\epsilon)-r)^2)((1+n\epsilon)-r)$$

$$= f'(r^2)r + \sum_{t_n=1+n\epsilon, \ n\geq 1} L_{t_n}(t_n-t_{n-1})\{f'((t_n-r)^2)(t_n-r) -$$

$$- f'((t_{n-1}-r)^2)(t_{n-1}-r)\}/(t_n-t_{n-1}) \tag{A.4}$$

where $L_t = |\{\beta; \ ||u^{(\beta)}-u^{(\alpha)}|| < t, \ \beta \neq \alpha, \ \beta \in A\}|$, $\epsilon > 0$ is an arbitrary constant. The first inequality in (A.4) comes from the Cauchy-Schwartz inequality, the second from the triangle inequality and the monotonicity of $f'(r^2)r$ w.r.t. r, and the third from the condition $||x-u^{(\alpha)}|| \leq r < 1$ and the monotonicity of $f'(r^2)r$. The lower limit on n is because of $L_t = 0$, $t \leq 1$, and the last equality holds due to this fact. Since $r < 1$, $f'((t-r)^2)(t-r)$ possesses a continuous derivative on $t \in [1,\infty)$, and L_t is measurable (since it is composed of a countable number of discrete steps), the r.h.s. of (A.4) is a continuous function of $\epsilon > 0$ which possesses the limit (as $\epsilon \to 0$): $f'(r^2)r + \int_1^{\infty} L_t \frac{d}{dt}\{f'((t-r)^2)(t-r)\}dt$, which is also a lower bound on the l.h.s. of (A.4), where in the derivation we assumed that this integral is finite, (i.e., at least $\lim_{t\to\infty} L_t \frac{d}{dt}\{f'((t-r)^2)(t-r)\} = 0$). In case it diverges, the same analysis can be done on $\{u^{(\alpha)}\}_{\alpha\in A}$ which are restricted to be in a sphere with radius ρ, which means $L_t = $ const. for $t > 2\rho$, and then the l.h.s. of (A.4) converges (for $\epsilon \to 0$) to

$$f'(r^2)r - L_{2\rho}f'((2\rho-r)^2)(2\rho-r) + \int_1^{2\rho} L_t \frac{d}{dt} f'((t-r)^2)(t-r)dt.$$

Thus, (13) will follow from the inequality $L_t \leq (2t+1)^N$, since $f'((t-r)^2)(t-r)$ is monotonically non-increasing.

However, this inequality follows from the condition $||u^{(\alpha)}-u^{(\beta)}|| \geq 1$, $\forall \alpha \neq \beta$, as the N dimensional sphere of radius $(t + \frac{1}{2})$, around $u^{(\alpha)}$, contains at least (L_t+1) disjoint spheres of radius $1/2$ each. Comparing the volumes of the large sphere and the (L_t+1) small ones we obtain the desired inequality.

(b)   Consider the case when the r.h.s. of (13) diverges. Then even if we consider this integral with lower limit $T \gg 1$ it still diverges to $+\infty$. A well known sphere packing result is that there exists $\{u^{(\alpha)}\}_{\alpha \in A}$ such that $\lim_{t \to \infty} [L_t/(2t+1)^N] \geq \delta > 0$ (c.f., [22]). For these $\{u^{(\alpha)}\}_{\alpha \in A}$, the last line in (A.4) can be arbitrarily large, negative numbers for small ε, for every $r > 0$, as $f'(r^2)r$ is finite. Furthermore, we can obtain this result also when $u^{(\alpha)}$ is at the origin and $u^{(\beta)}$, $\beta \neq \alpha$ are all at the upper half space (i.e., the first coordinate of $u^{(\beta)}$ is non-negative). Consider for any $r > 0$ the state x with first coordinate equals to r, and the rest being zero. As $t \to \infty$ the distribution of the $\{u^{(\alpha)}\}_{\alpha \in A}$ elements in an infinitesimal disk between the spheres of radius t and $(t + \Delta t)$ becomes spherically uniform in the upper half space; therefore, as

$$\lim_{t \to \infty} \left\{ \frac{1}{\Delta t (2t+1)^{N-1}} \int_{-\pi/2}^{\pi/2} \cos \theta \, dA \right\} > 0$$

(where dA is a volume element on this disk, and θ is the phase w.r.t. to the first coordinate axis), for the chosen x, $(\nabla V, x-u^{(\alpha)}) = -\infty$ provided that the second line in (A.4) diverges. However, we already know that the last line in (A.4) diverges, even when only $t \geq T \gg 1$ is considered, and for these values of $t = ||u^{(\beta)}-u^{(\alpha)}||$, $||u^{(\beta)}-u^{(\alpha)}|| \sim ||u^{(\beta)}-x|| + ||u^{(\alpha)}-x||$, as $r = ||x-u^{(\alpha)}|| \ll t$. Thus, both the second and the last lines of (A.4) diverge together.

To conclude, we have shown that there is a sphere packing construction with $u^{(\alpha)} = 0$, for which whenever the r.h.s. of (13) diverges, choosing x(0) with the first coordinate arbitrarily small positive, and the rest of them zero, will result in $\dot{x}$ with arbitrarily large positive first coordinate and the rest of them zero (using symmetry arguments), so that x(t) will move along the positive part of the first axis and never converges to $u^{(\alpha)} = 0$. Thus, $\varepsilon_{max} = 0$ in this case. $\square$

Proof of Lemma 3. By adding $1_{x \notin Q} g(||x-\bar{u}||^2)$ to $V(\cdot)$ we have not changed $V(\cdot)$ or the evolution (1) in the interior of Q. Thus, we only have to prove that there are no fixed points of (1) outside Int Q.

Assume that $x_0 \notin$ Int Q is a fixed point of (1), and denote by $C \subset$ Int Q the convex hull of $\{u^{(\alpha)}\}_{\alpha \in A} \cup \{\bar{u}\}$, then there is a convex, closed, neighborhood $U(x_0)$ of x_0, such that $U(x_0) \cap C = \Phi$ (as C is a closed set). Thus, there is an hyperplane $\mathscr{U}$ that strictly separates C and $U(x_0)$ (which is also compact); let n denote the vector normal to $\mathscr{U}$ towards C. Now, on $U(x_0)$:

$$(\dot{x}, n) = + \sum_{\alpha \in A} 2f'_\alpha(||x-u^{(\alpha)}||^2)(u^{(\alpha)}-x, n) + 1_{x \notin Q} 2g'(||x-\bar{u}||^2)(\bar{u}-x, n) > 0 \qquad (A.5)$$

where the inequality follows from the monotonicity of the $f_\alpha(\cdot)$'s and $g(\cdot)$, and the geometry of the problem. Thus, in particular $\dot{x} \neq 0$ at $x_0 \in U(x_0)$, which contradicts the assumption that x_0 is a fixed point of (1).

We have also shown by (A.5) that there is a drift towards C, from any point $x \notin C$. $\quad \square$

Proof of Lemma 4. Let us define $R(t) = ||x(t)-u^{(\alpha)}||$, then for evolution (1):

$$\dot{R}(t) = -\frac{1}{R(t)}(\nabla V(x(t)), x(t)-u^{(\alpha)})$$

$$\leq -2\left\{f'(r^2)r + \int_1^\infty \frac{d}{d\mu}[f'((\mu-r)^2)(\mu-r)]L_\mu d\mu\right\} \qquad (A.6)$$

where $r \triangleq \theta\hat{\epsilon}(m,N)$, and in deriving (A.6), we used (A.4), and the condition $||x(0) - u^{(\alpha)}|| \leq \theta\hat{\epsilon}(m,N)$ which ensures that $R(t) \leq R(0) \leq r$ (due to (15)). However, (13) - (15) also bound the r.h.s. of (A.6) for $f(d) = -k(\frac{d}{d_0})^{-m}$, and give:

$$\dot{R}(t) \leq -2kmd_0^m\left\{\left(\frac{1}{\theta\hat{\epsilon}(m,N)}\right)^{(2m+1)} - \left(\frac{1 - \hat{\epsilon}(m,N)}{\hat{\epsilon}(m,N)(1-\theta\hat{\epsilon}(m,N))}\right)^{(2m+1)}\right\}. \qquad (A.7)$$

Integrating (A.7), and using the fact that $R(t) \geq 0$, we obtain $R(t) \leq 0$ for $t \geq T$, where

$$T = \left(\frac{\theta\hat{\epsilon}(m,N)}{\sqrt{d_0}}\right)^{(2m+2)} \frac{d_0}{2km} \left\{1 - \left(\frac{\theta(1-\hat{\epsilon}(m,N))}{1-\theta\hat{\epsilon}(m,N)}\right)^{(2m+1)}\right\}^{-1}. \tag{A.8}$$

However, $R(t) \leq 0$ implies $R(t) = 0$, that implies $x(t) = u^{(\alpha)}$ for $t \geq T$. $\quad\square$

Proof of Theorem 2. Consider a neighborhood y of $x(n)$ with Hamming distance one, then either (A) $||y-u^{(i)}|| = ||x(n)-u^{(i)}|| - \frac{2}{N}$, or (B) $||y-u^{(i)}|| = ||x(n)-u^{(i)}|| + \frac{2}{N}$, and for $x(n) \neq u^{(i)}$, there are exactly $\frac{N}{2}||x(n)-u^{(i)}|| \geq 1$ neighbors of type (A). The theorem is thus a direct consequence of the following claim (when (17) holds):

Claim: For any x such that $||x-u^{(i)}|| \leq 2\theta\rho$ then $V(y) < V(x)$ for neighbors y of type (A), and $V(y) \geq V(x)$ for neighbors y of type (B).

Proof of the Claim. Note that for any $1 \leq j \leq K$, $-\frac{2}{N} \leq ||y-u^{(i)}|| - ||x-u^{(j)}|| \leq \frac{2}{N}$, since y and x differ only in one component. Furthermore, it is enough to show that $V(y) \geq V(x)$ for neighbors y of type (B), with strict inequality for $||x-u^{(i)}|| < 2\theta\rho$, since if y is of type (A) w.r.t. x, then $||y-u^{(i)}|| \leq 2\theta\rho$ as well, and x is of type (B) w.r.t. y.

Since the function $f(d)$ is monotonically increasing, and y is of type (B):

$$f(||y-u^{(j)}||^2) \geq f((||x-u^{(j)}|| - \frac{2}{N})^2) \qquad \forall j \neq i$$
$$f(||y-u^{(i)}||^2) = f((||x-u^{(i)}|| + \frac{2}{N})^2) \tag{A.9}$$

So,

$$V(y) - V(x) \geq \left\{(||x-u^{(i)}||)^{-2m} - (||x-u^{(i)}|| + \frac{2}{N})^{-2m}\right\}$$
$$- \sum_{\substack{j=1\\j\neq i}}^{K} \left\{(||x-u^{(j)}|| - \frac{2}{N})^{-2m} - (||x-u^{(j)}||)^{-2m}\right\}. \tag{A.10}$$

But, $||x-u^{(i)}|| \leq 2\theta\rho$ and $||x-u^{(j)}|| \geq ||u^{(i)}-u^{(j)}|| - ||x-u^{(i)}|| \geq 2\rho(1-\theta)$, and the function $g(r) \triangleq r^{-2m} - (r+\Delta)^{-2m}$ is a monotonically decreasing function, so from (A.10):

$$V(y) - V(x) \geq \left\{ (2\theta\rho)^{-2m} - (2\theta\rho + \tfrac{2}{N})^{-2m} \right\}$$

$$- (K-1)\left\{ (2\rho(1-\theta) - \tfrac{2}{N})^{-2m} - (2\rho(1-\theta))^{-2m} \right\} \qquad (A.11)$$

with strict inequality whenever $||x-u^{(i)}|| < 2\rho\theta$. To complete the proof, we just have to show that the r.h.s. of (A.11) is non-negative whenever (17) holds. This is easily shown by a simple rearrangement of (A.11) using $\theta \leq 1/2$ and $(1-\theta) \geq 1/2$.

□

REFERENCES

[1] J. J. Hopfield, "Neural Networks and Physical Systems with Emergent Collective Computational Abilities", Proc. Nat. Acad. Sci. U.S.A., Vol. 79 (1982), pp. 2554-2558.

[2] Y. S. Abu-Mostafa and J. St. Jacques, "Information Capacity of the Hopfield Model", IEEE Trans. on Inf. Theory, Vol. IT-31 (1985), pp. 461-464.

[3] R. J. McEliece et al., "The Capacity of the Hopfield Associative Memory", to appear in IEEE Trans. on Inf. Theory, 1987.

[4] D. Kleinfeld, "Sequential State Generation by Model Neural Networks", submitted to Proc. Nat. Acad. Sci. U.S.A.

[5] L. Personnaz, et al., "Information Storage and Retrival in Spin-Glass-Like Neural Networks", J. Physique Lett., 46 (1985), pp. L-359-2-365.

[6] A. Dembo, "On the Capacity of the Hopfield Memory", to be submitted to IEEE Trans. on Inf. Theory.

[7] D. J. Amit, H. Gutfreund and H. Sampolinsky, "Statistical Mechanics of Neural Networks Near Saturation", Phys. Rev. Letters 55, p. 1530 (1985), extended version submitted to Phys. Reviews.

[8] I. Kantor and H. Sampolinsky, "Associative Recall of Memory without Errors", submitted to Phys. Reviews.

[9] D. J. Amit, H. Gutfreund and H. Sampolinsky, Phys. Rev. A32, pg. 1007 (1985).

[10] G. L. Weisbush and F. Fogelman-Soulie, J. Physique Lett., 46, pg. 624 (1985).

[11] P. Baldi and S. Venkatesh, "Higher Order Spin Glasses and the Number of Stable Points", submitted to Phys. Rev. Letters.

[12] P. Baldi, "Neural Networks, Orientations of the Hypercube and Algebraic Threshold Functions", submitted to IEEE Trans. of Inf. Theory.

[13] J. J. Hopfield and D. W. Tank, "Computing with Neural Circuits: A Model", Science, Vol. 233, August 1986, pp. 625-632.

[14] E. A. Coddington and N. Levinson, Theory of Ordinary Differential Equations, McGraw-Hill, New York, 1955.

[15] R. J. McEliece, The Theory of Information Theory and Coding, Encyclopedia of Mathematics and Its Applications, Vol. e, Addison Wesley, 1977.

[16] W. Rudin, Real and Complex Analysis, McGraw-Hill, New York, 1974.

[17] L. N. Cooper, "Neuron Learning to Network Organization", In: J. C. Maxwell, The Sesquicentennial Symposium, M. S. Berger, Editor, North Holland, Amsterdam, 1984.

[18] A. Odlyzko, Private Communication.

END

10-87

DTIC